

Minimax optimization of entanglement witness operator for the quantification of three-qubit mixed-state entanglement

Sungguen Ryu, Seung-Sup B. Lee,* and H.-S. Sim

Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea

(Dated: March 8, 2013)

We develop a numerical approach for quantifying entanglement in mixed quantum states by convex-roof entanglement measures, based on the optimal entanglement witness operator and the minimax optimization method. Our new approach is applicable to general entanglement measures and states, and it is an efficient alternative to the conventional approach based upon the optimal pure-state decomposition. Compared with the conventional one, it has two important merits (i) that the global optimality of the solution is quantitatively verifiable, and (ii) that the optimization is considerably simplified by exploiting the common symmetry of the target state and measure. To demonstrate the merits, we quantify Greenberger-Horne-Zeilinger (GHZ) entanglement in a class of three-qubit full-rank mixed states composed of the GHZ state, the W state, and the white noise, the simplest mixtures of states with different genuine multipartite entanglement, which have not been quantified before this work. We discuss some general properties of the form of the optimal witness operator, and of the convex structure of mixed states, which are related to the symmetry and the rank of states.

PACS numbers: 03.67.Mn, 03.65.Ud

I. INTRODUCTION

Quantum entanglement, a quantum correlation among subsystems, is at the core of strange quantum phenomena, and it is the key ingredient of quantum information processing [1]. Its detection and quantification has attracted much attention of varieties of scientific communities for both fundamental and practical purposes.

Entanglement measure $\mathcal{E}(\rho)$ quantifies the entanglement contained in a quantum state ρ [2]. The definition of $\mathcal{E}(\rho)$ is not unique and depends on the type of entanglement to be quantified, as there exist many different types of entanglement. A measure is constructed under the restrictions by the physical meaning of entanglement: (i) Any measure vanishes for separable states $\rho = \sum_i p_i \rho_{A,i} \otimes \rho_{B,i} \otimes \cdots$, (ii) is invariant under local unitary operations, and (iii) does not increase under local quantum operations and classical communications (LOCC). The last property, so-called monotonicity under LOCC, is usually treated by adopting the convexity of a measure, $\mathcal{E}(p\rho_1 + (1-p)\rho_2) \leq p\mathcal{E}(\rho_1) + (1-p)\mathcal{E}(\rho_2)$ [3]. A convex-roof measure, the most popular type of measure, is defined first for pure states $|\psi\rangle$, and then extended, based on the convexity, to mixed states ρ via the convex roof construction

$$\mathcal{E}(\rho) = \inf_{\{p_i, \psi_i\}} \sum_i p_i \mathcal{E}(|\psi_i\rangle), \quad (1)$$

where the optimization (“inf”) runs over every possible pure state decomposition of $\rho = \sum_i p_i \pi_{\psi_i}$ and $\pi_{\psi} \equiv |\psi\rangle\langle\psi|$ denotes the pure-state projector.

In most cases, except for two-qubit states [4–6], the quantification of mixed-state entanglement requires

heavy computational cost for the optimization of a large number of parameters in Eq. (1) [7, 8]. The pure-state decomposition of a rank- n mixed state into m pure states utilizes a $m \times n$ left-unitary matrix which has $2mn - n^2 - 1$ parameters. As m ranges from n to n^2 according to Carathéodory’s theorem [9], the maximal number of parameters for the optimization becomes $\sim 2n^3$; it already reaches a thousand for three-qubit full-rank states. The optimization of such a large number of parameters is prone to the convergence to a local optimum, resulting in the overestimation of $\mathcal{E}(\rho)$. In usual optimization problems, it cannot be ensured, without the overall investigation of the parameter space, whether the result of the optimization is globally optimal, i.e. provides the right answer. These facts make the quantification of multipartite and/or high-dimensional mixed-state entanglement practically inaccessible theoretically and experimentally.

There exists another definition of $\mathcal{E}(\rho)$, which is equivalent to Eq. (1). It is the dual problem of Eq. (1) in the sense that it is based on the optimization of quantum operators rather than the optimization of states; see Eq. (2) below. The optimized operator, whose expectation value provides $\mathcal{E}(\rho)$, is called the optimal witness operator [10–12]; here the optimal witness operator is used for entanglement quantification rather than detection [13, 14]. This concept has been proved to be useful for estimating a lower bound of $\mathcal{E}(\rho)$ with partial information of ρ accessed in experiments [15].

Very recently, it was proposed [16] by two of the authors of the present work that an efficient way of constructing the optimal witness operator can relieve the difficulty of optimizing $\mathcal{E}(\rho)$. The benefits of this approach as an optimization problem are: (i) The size of the parameter space for this optimization is much smaller than that for Eq. (1), and (ii) one can verify whether the result of this approach is globally optimal, i.e. provides

* ssblee@kaist.ac.kr

the true answer, based on the duality between this approach and Eq. (1). This approach is generally applicable to any \mathcal{E} and ρ . This approach successfully quantifies three-qubit Greenberger-Horne-Zeilinger (GHZ) entanglement and four-qubit bound entanglement in some full-rank mixed states ρ that have been usually prepared in laboratories, even providing the analytic expression of $\mathcal{E}(\rho)$, demonstrating its usefulness [16]; this was, to our knowledge, the first quantification of entanglement in multipartite full-rank mixed states.

In this paper, we formulate how to numerically obtain the optimal witness operator, utilizing the minimax optimization method [17]; the utilization is shortly mentioned in Ref. [16], but its detail was not investigated. To apply the minimax method, we derive the canonical form of the witness operator that is generally applicable to any measures and states; see Theorem 1 below. We devise the techniques for handling the non-analyticity in the minimax method and of reinforcing the robustness against numerical errors, and also exploit the existing theories of the minimax method [17, 18] to enhance computation efficiency. We emphasize that in our approach, the global optimality of a witness operator is quantitatively verified by the minimal distance between the target state and the set of the states exactly quantified by the witness operator. We also notice that the common symmetries of the target state and measure simplify the optimization.

Our method is applied to scrutinize three-qubit GHZ entanglement in exemplar states. In particular, we quantify GHZ entanglement in a family of full-rank mixed states composed of the GHZ state, the W state, and the white noise, which is the simplest full-rank mixture of pure states with different genuine multipartite entanglement; this type of mixed states has not been quantified before, to our knowledge. This reveals some general properties of the form of the optimal witness operator, and of the convex structure of mixed states, which are related to the symmetry and the rank of the states.

This paper is organized as follows. In Sec. II, we introduce the optimal witness operator. In Sec. III, we develop the minimax optimization, how to verify the global optimality of its solution, and the simplification utilizing the symmetries. In Sec. IV, we apply our approach to examples, and discuss some general properties of the form of the optimal witness operator. We conclude in Sec. V.

II. OPTIMAL WITNESS OPERATOR

We sketch how to quantify mixed-state entanglement by the optimal witness operator, which was developed in Ref. [16]. The primal problem in Eq. (1) is equivalently rephrased as the dual problem [10, 16] that $\mathcal{E}(\rho)$ is obtained by the optimal operator X_ρ ,

$$\begin{aligned} \mathcal{E}(\rho) &= \text{Tr}(X_\rho \rho) = \sup_{X \in \mathbb{M}_{\mathcal{H}}} \text{Tr}(X \rho), \\ \mathbb{M}_{\mathcal{H}} &\equiv \{X \mid \text{Tr}(X \pi_\psi) \leq \mathcal{E}(|\psi\rangle) \quad \forall |\psi\rangle \in \mathcal{H}\}. \end{aligned} \quad (2)$$

Here \mathcal{H} is a Hilbert space that includes the range of ρ , and any Hermitian operator X satisfies the non-overestimating condition of $\text{Tr}(X \rho_1) \leq \mathcal{E}(\rho_1) \quad \forall \rho_1 \in \mathcal{H}$ so that $\text{Tr}(X \rho)$ is a strict lower bound of $\mathcal{E}(\rho)$. X behaves as a *generalized* entanglement witness operator: In the convex structure of quantum states, the hyperplane of ρ_0 satisfying $\text{Tr}(X \rho_0) = \mathcal{E}_0$ isolates the convex set of $\{\rho_1 \mid \mathcal{E}(\rho_1) \leq \mathcal{E}_0\}$, namely, $\text{Tr}(X \rho_1) \leq \mathcal{E}_0$ for any ρ_1 with $\mathcal{E}(\rho_1) \leq \mathcal{E}_0$. Hence X “witnesses” entanglement of finite amount \mathcal{E}_0 , as $\text{Tr}(X \rho) > \mathcal{E}_0$ implies that $\mathcal{E}(\rho) > \mathcal{E}_0$. Note that the conventional entanglement witness [14] detects entanglement as it isolates the set of separable states with $\mathcal{E}_0 = 0$; for mathematical simplicity, in Eq. (2), we use the different convention of a witness operator by the sign factor of -1 from the definition in literature.

The optimal entanglement witness operator [10–12, 16] X_ρ satisfies $\text{Tr}(X_\rho \rho) = \mathcal{E}(\rho)$, quantifying entanglement. It depends on the target state ρ and the choice of \mathcal{E} . From the viewpoint of optimization theory, $X = X_\rho$ is the globally optimal point in $\mathbb{M}_{\mathcal{H}}$ that maximizes $\text{Tr}(X \rho)$. The global optimum is not necessarily unique for a given ρ , and any single global optimum suffices to obtain $\mathcal{E}(\rho)$. In some cases, X_ρ does not exist in $\mathbb{M}_{\mathcal{H}}$, and the supremum in Eq. (2) is only asymptotically accessed by an asymptotic form of a witness operator [16]. The existence of X_ρ depends on the choice of \mathcal{H} .

We mention one important finding of Ref. [16]. The duality between Eqs. (1) and (2) affords the criterion that is used to verify the global optimality of a solution of Eq. (2). From the fact that $\sum_i p_i \langle \psi_i | X | \psi_i \rangle \leq \mathcal{E}(\rho) \leq \sum_i p_i \mathcal{E}(|\psi_i\rangle)$, where $\{p_i, |\psi_i\rangle\}$ is a pure state decomposition of $\rho = \sum_i p_i \pi_{\psi_i}$, we have the following equivalence: If X is the optimal witness operator of ρ , ρ belongs to the convex hull of \mathbb{P}_X , $\text{conv } \mathbb{P}_X$. The converse is also true,

$$X = X_\rho \Leftrightarrow \rho \in \text{conv } \mathbb{P}_X \quad (3)$$

where \mathbb{P}_X is the pure-state set of $\{|\psi\rangle \mid \text{Tr}(X \pi_\psi) = \mathcal{E}(|\psi\rangle)\}$. This implies that we can verify whether a witness operator X is the optimal one for ρ , by checking whether ρ is decomposed into a convex combination of the elements of \mathbb{P}_X . Hence, it is possible to check whether a numerical output of the optimization of Eq. (2) is at a local optimum (giving only a lower bound of \mathcal{E}) or at the global optimum (with the exact value of \mathcal{E}). We provide a quantitative way of checking this in Sec. III C.

III. MINIMAX OPTIMIZATION OF WITNESS

This section contains the followings. In Sec. III A, we formulate the minimax optimization of X_ρ . In Sec. III B, we devise new techniques to overcome the difficulties that arise in the minimax optimization. In Sec. III C, based upon the correspondence of Eq. (3), we give how to quantitatively check whether the optimization result is globally optimal. In Sec. III D, we show that our optimization is simplified when \mathcal{E} and ρ have common symmetries. We state the algorithm for the optimization in Sec. III E.

A. Formulation

To obtain X_ρ , one optimizes the parameters of X under the condition of $\text{Tr}(X\pi_\psi) \leq \mathcal{E}(|\psi\rangle)$ in Eq. (2). To handle this condition efficiently, we translate it into another optimization problem. This idea of “optimization of optimization” leads to the essential part of our theory:

Theorem 1. *The optimization of the optimal witness operator X_ρ in Eq. (2) is formulated as*

$$\begin{aligned} \mathcal{E}(\rho) &= \text{Tr}(X_\rho \rho) = \sup_{\Pi \geq 0} \inf_{|\psi\rangle \in \mathcal{H}} F(\Pi, |\psi\rangle) = \sup_{\Pi \geq 0} G(\Pi), \\ F(\Pi, |\psi\rangle) &\equiv \text{Tr}((\Pi - \mu I)\rho), \\ \mu(\Pi, |\psi\rangle) &\equiv \text{Tr}(\Pi\pi_\psi) - \mathcal{E}(|\psi\rangle), \\ G(\Pi) &\equiv \text{Tr}((\Pi - \mu_\Pi I)\rho), \quad \mu_\Pi \equiv \sup_{|\psi\rangle \in \mathcal{H}} \mu(\Pi, |\psi\rangle). \end{aligned} \quad (4)$$

Proof. We decompose each Hermitian operator $X \in \mathcal{H}$ into $X = \Pi - \mu I$ with $\Pi \geq 0$. Since X does not overestimate \mathcal{E} , μ is lower-bounded by $\mu \geq \mu_\Pi$. For a given Π , the finest witness operator is $X = \Pi - \mu_\Pi I$ since a smaller μ gives a finer X . Then, the expectation value $\text{Tr}X\rho$ is optimized by varying Π , leading to Eq. (4). \square

The minimization (“inf”) directly reflects the non-overestimating condition in Eq. (2). We note that X_ρ has an asymptotic form, when the supremum in Eq. (4) is asymptotically accessed; in this access, $\text{Tr}((\Pi - \mu_\Pi I)\rho)$ converges, while Π and μ_Π diverge.

The optimization in Eq. (4) is a kind of the linear minimax problem [17]. The minimax problem has been widely applied to n -person games, finance, economics, and policy optimization. Since it is an optimization problem (“sup”) that includes a sub-optimization (“inf”), it is non-trivial to solve due to the following difficulty. In a simple-minded approach, one minimizes $F(\Pi, |\psi\rangle)$ over $|\psi\rangle$ at every fixed Π , and then maximizes the resulting $G(\Pi)$ over Π . To apply usual competitive optimization algorithms to the maximization of $G(\Pi)$, one needs to evaluate the first or the higher order derivatives of $G(\Pi)$. The evaluation of the derivatives requires auxiliary minimization to attain $G(\Pi + d\Pi)$ for small deviation $d\Pi$ in every direction in the domain of Π . Then, the computational cost increases enormously as the dimension of the parameter space grows.

Fortunately, there exists an efficient strategy that reduces the computational cost in evaluating the derivatives of $G(\Pi)$. We here sketch the strategy following Ref. [17]. First we define the solution set $\mathbb{R}(\Pi)$

$$\mathbb{R}(\Pi) \equiv \{|\psi\rangle \in \mathcal{H} \mid F(\Pi, |\psi\rangle) = G(\Pi)\}, \quad (5)$$

which becomes equivalent to \mathbb{P}_X when $X = \Pi - \mu_\Pi I$. Namely, the elements in $\mathbb{R}(\Pi)$ provide the value of G at Π . The set \mathbb{R} plays a key role in obtaining the derivatives of G ; it is also useful for checking the global optimality of the solution of Eq. (4) as shown in Sec. III C. The

behavior of G near Π is governed by $\mathbb{R}(\Pi)$; for small deviation $d\Pi$, $G(\Pi + d\Pi) = \inf_{|\psi\rangle \in \mathbb{R}(\Pi)} F(\Pi + d\Pi, |\psi\rangle)$.

When \mathbb{R} has multiple elements, the multivariate function G is not differentiable, i.e. the derivative of G cannot be expressed by well-defined Jacobian or gradient. For this case, we use $D_{\delta\Pi}G(\Pi)$, the directional derivative of G along $\delta\Pi$ at Π , and $D_{\hat{\Pi}}G(\Pi)$, that along the direction of steepest ascent $\delta\hat{\Pi}$ of $G(\Pi)$,

$$\begin{aligned} D_{\delta\Pi}G(\Pi) &= \inf_{|\psi\rangle \in \mathbb{R}(\Pi)} \text{Tr}[(\rho - \pi_\psi \text{Tr}\rho)\delta\Pi], \\ D_{\hat{\Pi}}G(\Pi) &= \sup_{\substack{\delta\Pi \in \Gamma(\Pi) \\ \|\delta\Pi\|=1}} D_{\delta\Pi}G(\Pi) \end{aligned} \quad (6)$$

where $\Gamma(\Pi)$ is the cone of possible directions of $\delta\Pi$ at Π , and the directions are normalized, $\|\delta\Pi\| = \|\delta\hat{\Pi}\| = 1$, by the norm of $\|M\| = \sqrt{\text{Tr}M^\dagger M}$.

Utilizing Eq. (6), $G(\Pi)$ can be maximized by the steepest descent method. However, there are still two difficulties (i) that the direct evaluation of the steepest ascent using Eq. (6) is another minimax problem, which may be laborious, and (ii) that the Hessian (the second order derivative) of $G(\Pi)$ may be unavailable when $G(\Pi)$ is not differentiable (i.e. when $\mathbb{R}(\Pi)$ has multiple elements). Note that the Hessian is used in most of efficient optimization algorithms since it informs the local behavior of the function to be optimized, which cannot be obtained from the gradient. For example, the Hessian determines whether a locally optimal solution is stable or unstable by its eigenvalues. We below devise new techniques to overcome the above difficulties.

B. Minimization of $F(\Pi, |\psi\rangle)$

To solve the minimax problem in Eq. (4), we need to accurately solve its “inf” problem, avoiding the convergence to local optima. Below, we propose the techniques that improve the optimization performance in the inf problem. We emphasize that the techniques are applicable to general minimax problems.

Parallelization.— The inf problem can be solved by adapting a parallelization approach with multiple solvers. A solver means a single optimization procedure that finds an (local or global) optimum, following a certain algorithm and starting from an initial guess (randomized or not) of the solution. We collect the results of the solvers, and assign the best value among them to G . Then $\mathbb{R}(\Pi)$ is comprised of $|\psi\rangle$ ’s providing the best value or the sufficiently good values slightly deviated from the best one within a given tolerance. $\mathbb{R}(\Pi)$ can be fully identified, in principle, given that the unlimited number of solvers are available. In practical situations, however, parallelization details such as the number of solvers and initial guesses need to be tuned depending on the problem.

To design efficient parallelization, one may require to investigate entanglement classes with respect to a given

measure \mathcal{E} . The entanglement class is an exclusive subset of quantum states whose elements share the common type of entanglement. For example, there are two entanglement classes for two qubits: one is entangled and the other is separable. When confined to pure states, some classes, e.g. two-qubit separable states, have vanishing volume. It is generally possible that some elements of $\mathbb{R}(\Pi)$ belong to different entanglement classes, even to classes with zero volume. However, if a solver is designed to minimize $F(\Pi, |\psi\rangle)$ over the full set of pure states at once, the zero-volume classes can be rarely accessed and the solution tends to appear in classes with finite volume. This can lead to the insufficient construction of $\mathbb{R}(\Pi)$.

This problem is resolved by the divide-and-conquer approach: (i) divide the full set of pure states into the entanglement classes with respect to the target entanglement \mathcal{E} , (ii) assign solvers to different classes, and then (iii) let each individual solver minimize $F(\Pi, |\psi\rangle)$ over $|\psi\rangle$ only within the assigned class. In this approach, all the classes can be thoroughly accessed. Though the full classification of entanglement for more than three qubits is complicated, even an incomplete classification, such as the dichotomy between finite and vanishing \mathcal{E} , improves the performance of the optimization.

Finding elements of $\mathbb{R}(\Pi)$.— We need sufficient information of \mathbb{R} , to precisely evaluate the directional derivative of G or to verify the global optimality of the optimization result. \mathbb{R} may be numerically found by considering the local optima of the inf problem, as mentioned above. This procedure has two non-trivial points: (i) The values of F at the local optima may deviate due to numerical errors, and (ii) all local optima residing in the high-dimensional domain are hard to be found.

We handle the point (i) by extending the concept of the (exact) solution set \mathbb{R} to the set $\tilde{\mathbb{R}}$ of the candidates for the solutions. $\tilde{\mathbb{R}}$ consists of not only the global optima but also the local optima which give the value of F sufficiently close to the globally optimal one. This extension makes G more robust against errors in numerical computation, avoiding the possibility that the global optima are treated, due to the errors, as local optima and excluded from \mathbb{R} ; the exclusion will induce the failure in evaluating $D_{\delta\Pi}G$, hence the maximization of G .

On the point (ii), the parallelization of the inf problem is helpful, as it provides a number of local optima, the elements of $\tilde{\mathbb{R}}$. Sometimes, more elements of $\tilde{\mathbb{R}}$ are required. They are provided by the local maximum points $|\psi\rangle \in \mathcal{H}$ of the function

$$d(|\psi\rangle, |\psi_0\rangle) - k(F(\Pi, |\psi\rangle) - F(\Pi, |\psi_0\rangle)), \quad (7)$$

where $d(|\psi\rangle, |\psi_0\rangle)$ measures the distance between $|\psi\rangle$ and the best optima $|\psi_0\rangle$ obtained from the parallelization, and k is a positive number fixed arbitrarily large. The exploration of the local maximum points by Eq. (7) depicts the search of $|\psi\rangle$'s that are located far from the best known optimum $|\psi_0\rangle$ but provide the value of F close to the best one; the second term in the cost function of Eq. (7) acts as the barrier that rules out the points $|\psi\rangle$

providing not-small-enough values of F . With reasonable tolerance for F values, the above search accesses all available candidates of $\tilde{\mathbb{R}}$. The search can be generalized by considering also the distance from the other locally optimal candidates in $\tilde{\mathbb{R}}$. Note that if a better optimum point giving a smaller value of F than $|\psi_0\rangle$ appears in this search, then $|\psi_0\rangle$ should be replaced with this better one. This search complements the parallelization in finding the global optima of the inf problem.

Smoothing of $G(\Pi)$.— The function G resulting from the inf problem can be non-analytic, as mentioned in Sec. III A. To surmount the non-analyticity and to efficiently evaluate the derivatives of G , we approximate G to an analytic function \tilde{G} within good accuracy. We call this approximation the “smoothing” of G .

In this method, we use the candidate set $\tilde{\mathbb{R}} = \{|\psi_1\rangle, \dots, |\psi_N\rangle\}$ instead of \mathbb{R} , as the computation of G becomes more robust against numerical errors. When $\tilde{\mathbb{R}}$ has enough elements, G is faithfully approximated by \tilde{G} ,

$$\tilde{G}(\Pi) \equiv H(F(\Pi, |\psi_1\rangle), \dots, F(\Pi, |\psi_N\rangle); b). \quad (8)$$

$H(\alpha_1, \alpha_2, \dots, \alpha_N; b)$ is an analytic function imitating $\min\{\alpha_n\}$ with a smoothing parameter $b > 0$, satisfying

$$H(\alpha_1, \dots, \alpha_N; b) \leq \lim_{b \rightarrow 0} H(\alpha_1, \dots, \alpha_N; b) = \min_{n=1, \dots, N} \alpha_n.$$

Larger b makes H smoother, but more deviated from $\min_{n=1, \dots, N} \{\alpha_n\}$. As $\tilde{G}(\Pi)$ is a differentiable function with gradient $\nabla \tilde{G}$, the directional derivative in Eq. (6) is approximately obtained by a simplified form

$$D_{\delta\Pi} \tilde{G}(\Pi) = \text{Tr}(\nabla \tilde{G} \delta\Pi), \quad (9)$$

$$\nabla \tilde{G} = \frac{\partial H(\Pi, \{|\psi_i\rangle\})}{\partial \Pi}.$$

Similarly, the direction of the steepest ascent is given by $\nabla \tilde{G} / \|\nabla \tilde{G}\|$. Accordingly, we employ the gradient and the Hessian to maximize \tilde{G} efficiently and precisely, with avoiding the difficulty in directly evaluating Eq. (6).

We mention our choice of the smoothing function H ,

$$H(\alpha_1, \alpha_2; b) \equiv \frac{\alpha_1 + \alpha_2}{2} - \sqrt{b^2 + \frac{(\alpha_1 - \alpha_2)^2}{4}},$$

$$H(\alpha_1, \dots, \alpha_N; b) \equiv H(H(\alpha_1, \dots, \alpha_{N-1}; b), \alpha_N; b).$$

For $N \geq 3$, H is not exactly symmetric under the permutation of α_n , but, the asymmetry is negligible when b is small enough. Then Eqs. (8) and (9) are rewritten as

$$\tilde{G}(\Pi) = \text{Tr}(\Pi \rho) - \tilde{\mu}_\Pi,$$

$$\tilde{\mu}_\Pi \equiv -H(-\mu(\Pi, |\psi_1\rangle), \dots, -\mu(\Pi, |\psi_N\rangle); b) \geq \mu_\Pi,$$

$$\nabla \tilde{G} = \rho - \frac{\partial \tilde{\mu}_\Pi}{\partial \Pi}.$$

Here, we have used the property of $H(\alpha_1 + \alpha', \dots, \alpha_N + \alpha'; b) = H(\alpha_1, \dots, \alpha_N; b) + \alpha'$.

C. Verification of global optimality

The maximization of G in Eq. (4) provides a candidate of the optimal witness operator. We will show that in our approach, one can verify whether the candidate is the global optimum of the problem, i.e. the correct optimal witness operator.

We first translate the duality correspondence in Eq. (3) into the following, to use it in the minimax approach.

Theorem 2. *A witness operator $X = \Pi - \mu_\Pi I$ is optimal for $\mathcal{E}(\rho)$ if and only if $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi)) = 0$. Here, $d_{\min}(\rho'; \mathbb{S}) = \min_{\sigma \in \mathbb{S}} \|\rho' - \sigma\|_{\text{HS}}$ measures the minimal distance between a state ρ' and a set \mathbb{S} of states σ by the Hilbert-Schmidt norm $\|\cdots\|_{\text{HS}}$.*

Proof. \mathbb{P}_X is equivalent to $\mathbb{R}(\Pi)$ when $X = \Pi - \mu_\Pi I$. And $\rho \in \text{conv } \mathbb{P}_X$ if and only if $d_{\min}(\rho; \text{conv } \mathbb{P}_X) = 0$. \square

Based on Theorem 2, we use $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ for quantitatively judging how far a candidate Π for the solution of Eq. (4) is deviated from the global optimum. Concretely, $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ is not a distance between the candidate and the global optimum, but a degree of the deviation from the duality correspondence in Eq. (3) which should be satisfied by the global optimum. Hence any *a priori* information about the global optimum is not necessary, and the only requirement is that the elements of $\mathbb{R}(\Pi)$ are sufficiently found by the parallelization of the inf problem or by the exploration using Eq. (7).

To use Theorem 2 in numerical optimization, the extension of \mathbb{R} to $\tilde{\mathbb{R}}$ and that of $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ to $d_{\min}(\rho; \text{conv } \tilde{\mathbb{R}}(\Pi))$ are crucial. When some elements of \mathbb{R} (therefore those of $\text{conv } \mathbb{R}(\Pi)$) are accidentally excluded due to numerical error, $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ can be overestimated even if Π is the globally optimal one. On the other hand, given an appropriate value of the tolerance for $F(\Pi, |\psi\rangle) - G(\Pi)$, $\mathbb{R}(\Pi)$ is the set of all relevant $|\psi\rangle$'s. The tolerance judges whether a state $|\psi\rangle$ is relevant to $\tilde{\mathbb{R}}(\Pi)$ by considering how accurately $\mathcal{E}(|\psi\rangle)$ is quantified by $X = \Pi - \mu_\Pi I$, since $\mathcal{E}(|\psi\rangle) - \text{Tr}(X\pi_\psi) = F(\Pi, |\psi\rangle) - G(\Pi)$ because of $\text{Tr}\pi_\psi = \text{Tr}\rho = 1$.

The choice of the tolerance for $F(\Pi, |\psi\rangle) - G(\Pi)$ governs the credibility of $d_{\min}(\rho; \text{conv } \tilde{\mathbb{R}}(\Pi))$. Too small tolerance is not helpful, as it causes the problem similar to the case of \mathbb{R} . Too large tolerance causes the inclusion of irrelevant elements $|\psi\rangle$'s in $\text{conv } \tilde{\mathbb{R}}(\Pi)$, for which $\text{Tr}(X\pi_\psi)$ is far deviated from $\mathcal{E}(|\psi\rangle)$. It leads to the underestimation of $d_{\min}(\rho; \text{conv } \tilde{\mathbb{R}}(\Pi))$, resulting in misleading verification of the global optimality. In our examples in Sec. IV, the values of $F(\Pi, |\psi\rangle)$ at $|\psi\rangle \in \tilde{\mathbb{R}}(\Pi)$ are well-separated from the values at other local optima, hence the adequate value of the tolerance is straightforwardly chosen.

The result of $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi)) = 0$ does not only confirm the global optimality of witness $X = \Pi - \mu_\Pi I$, but also provide the optimal pure-state decomposition of ρ for \mathcal{E} ; a convex sum $\in \text{conv } \mathbb{R}(\Pi)$ is the optimal decomposition of ρ , when it is identical to ρ and has the

vanishing minimal distance of $d_{\min} = 0$. It is because any convex sum $\rho' \in \text{conv } \mathbb{R}(\Pi) = \text{conv } \mathbb{P}_X$ itself is the optimal pure-state decomposition by Theorem 2 in Ref. [16].

We note the possibility that at some Π 's, $G(\Pi)$'s are similar to each other, but $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$'s are largely different; see the example in Fig. 2. This indicates that $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ is a more useful tool for checking the global optimality than the mere comparison of $G(\Pi)$'s.

D. Simplification by symmetry

The common symmetries of \mathcal{E} and ρ simplify the minimax approach. It is identified by the symmetry group $\mathbb{G} \equiv \{U | U^\dagger = U^{-1}, \rho = U\rho U^\dagger, \mathcal{E}(\rho') = \mathcal{E}(U\rho'U^\dagger) \ \forall \rho' \in \mathcal{H}\}$. Usually \mathbb{G} has local unitary operations and permutations of subsystems; see Sec. IV.

We first introduce some notions for later use. We call $O \in \mathcal{H}$ a symmetric operator with respect to \mathbb{G} when $O = UOU^\dagger$ holds $\forall U \in \mathbb{G}$. We also call $\mathbb{C} = \{P_i \in \mathcal{H}\}$ a “basis” set of symmetric operators P_i with respect to \mathbb{G} , when \mathbb{C} consists of the maximal number of linearly independent P_i 's satisfying $\text{Tr}(P_i P_j) = \delta_{ij}$. Any symmetric operator O is decomposed as a linear combination of P_i 's, and the coefficients of the decomposition can serve as the parametrization of O . And, any operator $O \in \mathcal{H}$ is symmetrized as $O^S = \sum_i \text{Tr}(O P_i) P_i = \sum_{U \in \mathbb{G}} U O U^\dagger$; it is obvious that $O^S = O$ if and only if O is symmetric. Note that though the choice of \mathbb{C} may not be unique, any choice provides the same result of the symmetrization.

In addition, we call a set \mathbb{S} a symmetric set with respect to \mathbb{G} when $\mathbb{S} = \{U O U^\dagger | \forall O \in \mathbb{S}, \forall U \in \mathbb{G}\}$; accordingly, a set of symmetric operators is symmetric. There exists a minimal subset \mathbb{S}^A of a symmetric set \mathbb{S} such that \mathbb{S} is generated from \mathbb{S}^A by applying the elements of \mathbb{G} to the elements of \mathbb{S}^A . We call \mathbb{S}^A the asymmetric unit of \mathbb{S} . It contains the information of \mathbb{S} not related to the symmetries. The choice of \mathbb{S}^A may not be unique, but any choice generates the same \mathbb{S} .

Based upon the notions, we find the following features:

Lemma 1. *There exist the optimal witness operator X_ρ and the corresponding Π_ρ symmetric with respect to \mathbb{G} .*

Proof. Consider the operator of $X = U^\dagger X_\rho U$ for an optimal witness operator X_ρ and any $U \in \mathbb{G}$. Then X is a witness operator $X \in \mathbb{M}_{\mathcal{H}}$ not overestimating \mathcal{E} since $\text{Tr}(U^\dagger X_\rho U \pi_\psi) = \text{Tr}(X_\rho U \pi_\psi U^\dagger) \leq \mathcal{E}(U|\psi\rangle) = \mathcal{E}(|\psi\rangle)$ holds $\forall U \in \mathbb{G}$ and $\forall |\psi\rangle \in \mathcal{H}$. The witness X is optimal for ρ , since $\mathcal{E}(\rho) = \text{Tr}(X_\rho \rho) = \text{Tr}(X \rho)$. Hence the symmetrization $X_\rho^S = \sum_i \text{Tr}(X_\rho P_i) P_i$ of a non-symmetric optimal witness operator X_ρ constructs the optimal witness operator X_ρ^S symmetric under $\forall U \in \mathbb{G}$. When X_ρ is symmetric, so is Π_ρ because I is symmetric. \square

Lemma 2. *If X and Π are symmetric with respect to \mathbb{G} , \mathbb{P}_X and $\mathbb{R}(\Pi)$ are also symmetric.*

Proof. $\forall |\psi\rangle \in \mathbb{P}_X$ and $\forall U \in \mathbb{G}$, $\mathcal{E}(|\psi\rangle) = \text{Tr}(X\pi_\psi) = \mathcal{E}(U|\psi\rangle) = \text{Tr}(U^\dagger XU\pi_\psi)$. This shows that $U|\psi\rangle \in \mathbb{P}_X$, indicating that \mathbb{P}_X is symmetric. And $\mathbb{R}(\Pi)$ is equivalent to \mathbb{P}_X for $X = \Pi - \mu_\Pi I$. \square

From these two Lemmas, we derive the simplification of the minimax optimization of the optimal witness operator, which is the main result of this subsection:

Theorem 3. *Suppose \mathcal{E} and ρ have the common symmetry given by \mathbb{G} . Then in solving Eq. (4), the domain of the sup problem $\Pi \geq 0$ and the solution set of the inf problem $\mathbb{R}(\Pi)$ can be replaced by $\Pi^S \geq 0$ and $[\mathbb{R}(\Pi^S)]^A$ (with respect to \mathbb{G}), respectively.*

Proof. We prove this theorem by the following two parts.

(i) $\Pi \rightarrow \Pi^S$: Lemma 1 ensures the existence of Π_ρ symmetric with respect to \mathbb{G} . Hence the optimum $\mathcal{E}(\rho)$ is accessible in the sub-domain of symmetric operators $\Pi^S \geq 0$. It proves the replacement of Π by Π^S .

(ii) $\mathbb{R}(\Pi) \rightarrow [\mathbb{R}(\Pi^S)]^A$: We start with $\mathbb{R}(\Pi^S)$ that is symmetric with respect to \mathbb{G} , replacing Π by Π^S as in the part (i). We consider a pure state $|\psi\rangle \in \mathbb{R}(\Pi^S)$. With small enough $r > 0$, one defines $B_r(|\psi\rangle)$, an open ball of radius r centered at $|\psi\rangle$; here the set of pure states is treated as a metric space with a Hilbert-Schmidt norm metric. Then for any $U \in \mathbb{G}$, the set of $\{U|b\rangle \mid \forall |b\rangle \in B_r(|\psi\rangle)\}$ becomes $B_r(U|\psi\rangle)$. Therefore $\mathbb{B} \equiv B_r(|\psi_1\rangle) \cup B_r(|\psi_2\rangle) \cup \dots$ for $\mathbb{R}(\Pi^S) = \{|\psi_1\rangle, |\psi_2\rangle, \dots\}$ is symmetric with respect to \mathbb{G} . Due to the symmetry, $\{F(\Pi^S, |\psi\rangle) \mid |\psi\rangle \in \mathbb{B}\} = \{F(\Pi^S, |\psi\rangle) \mid |\psi\rangle \in \mathbb{B}^A\}$ holds. It implies that the behavior of G near Π^S is fully described only by $[\mathbb{R}(\Pi^S)]^A$. \square

The replacement in Theorem 3 reduces the computation cost, since a symmetric operator is parametrized by a smaller number of parameters than a non-symmetric one. Namely, the dimension of the domains for the optimization is reduced by the symmetry. This reduction makes the optimization easier to converge to the global optima, since local optima (which are deep enough to be stuck) tend to appear less often in a lower dimensional space.

The minimal distance in Theorem 2, the quantitative criterion to verify the global optimality of a witness operator, is also streamlined by exploiting the symmetry, i.e. by replacing $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ to $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi^S))$.

We below apply the symmetries to numerical procedures. Using a “basis” set $\mathbb{C} = \{P_i\}$, we map operators in \mathcal{H} to real vectors in Euclidean space. Substituting $\Pi^S = \sum_i v_i P_i$ and $\rho = \sum_i r_i P_i$ into Eq. (4), we get

$$\mathcal{E}(\rho) = \sup_{\Pi \geq 0} \inf_{|\psi\rangle \in \mathcal{H}} \langle \mathbf{v}, \mathbf{r} \rangle - (\langle \mathbf{v}, \mathbf{q} \rangle - \mathcal{E}(|\psi\rangle)) \text{Tr} \rho,$$

where $\langle \cdot, \cdot \rangle$ is an inner product of vectors, $[\mathbf{v}]_i = v_i$, $[\mathbf{r}]_i = r_i$, and $[\mathbf{q}]_i = \text{Tr}(P_i \pi_\psi)$. Rewriting $F(\Pi, |\psi\rangle) = F(\mathbf{v}, \mathbf{q})$ and $G(\Pi) = G(\mathbf{v})$, the directional derivative of G at \mathbf{v} in direction \mathbf{v}' ($\|\mathbf{v}'\| = 1$) becomes [see Eq. (6)]

$$D_{\mathbf{v}'} G = \inf_{|\psi\rangle \in [\mathbb{R}(\Pi^S)]^A} \langle \mathbf{r} - \text{Tr} \rho \mathbf{q}, \mathbf{v}' \rangle.$$

To smooth G , we find the elements of $[\tilde{\mathbb{R}}(\Pi^S)]^A$ as the local maximum points $|\psi\rangle \in \mathcal{H}$ of the function in Eq. (7),

$$\|\mathbf{q} - \mathbf{q}_0\| + k(\mu(\Pi^S, |\psi\rangle) - \mu(\Pi^S, |\psi_0\rangle)),$$

where $\mu(\Pi^S, |\psi_0\rangle) = \mu_{\Pi^S}$, $[\mathbf{q}_0]_i = \text{Tr}(P_i \pi_{\psi_0})$, and k is a large positive constant. Then the gradient of the smoothed \tilde{G} [see Eq. (9)] is

$$\nabla \tilde{G} = \mathbf{r} + \frac{\partial H(-\mu(\mathbf{v}, \mathbf{q}_1), \dots, -\mu(\mathbf{v}, \mathbf{q}_N); b)}{\partial \mathbf{v}},$$

where $\mu(\mathbf{v}, \mathbf{q}_i) = \mu(\Pi, |\psi_i\rangle) = \langle \mathbf{v}, \mathbf{q}_i \rangle - \mathcal{E}(|\psi_i\rangle)$ for $[\mathbf{q}_i]_j = \text{Tr}(P_j \pi_{\psi_i})$ and $[\tilde{\mathbb{R}}(\Pi^S)]^A = \{|\psi_1\rangle, \dots, |\psi_N\rangle\}$.

The minimal distance is also rewritten from the Hilbert-Schmidt norm to the Euclidean norm of $d_{\min}(\mathbf{r}; \mathbb{S}) = \min_{\mathbf{s} \in \mathbb{S}} \|\mathbf{r}' - \mathbf{s}\|$. Then the sufficient and necessary condition of the global optimality becomes

$$d_{\min}(\mathbf{r}; \text{conv } \{\mathbf{q}_1, \dots, \mathbf{q}_N\}) = 0, \quad (10)$$

where $[\mathbf{r}]_i = \text{Tr}(P_i \rho)$, $[\mathbf{q}_i]_j = \text{Tr}(P_j \pi_{\psi_i})$, and $\{|\psi_1\rangle, \dots, |\psi_N\rangle\} = [\mathbb{R}(\Pi^S)]^A$ for a non-smoothed G and $\dots = [\tilde{\mathbb{R}}(\Pi^S)]^A$ for the smoothed \tilde{G} .

By finding $\mathbf{r}' = \sum_i w'_i \mathbf{q}_i \in \text{conv } \{\mathbf{q}_1, \dots, \mathbf{q}_N\}$ that provides $d_{\min} = 0$, we build the optimal pure-state decomposition $\rho = \sum_i w'_i \pi_{\psi_i}^S$, where $\pi_{\psi_i}^S = \sum_{U \in \mathbb{G}} U \pi_{\psi_i} U^\dagger$. We note that $[\mathbb{R}(\Pi^S)]^A$ or $[\tilde{\mathbb{R}}(\Pi^S)]^A$ holds the key information of the convex structure of $\mathbb{R}(\Pi^S)$ with respect to \mathcal{E} , hence, is the minimal set for analysis, since $\mathbb{R}(\Pi^S)$ is constructed from $[\mathbb{R}(\Pi^S)]^A$ by the symmetry operations.

E. Algorithm

We summarize this section outlining the entire algorithm for the minimax optimization of witness operator for a general state ρ and a measure \mathcal{E} . The steps of the algorithm are as follows.

1. Investigate the common symmetry of \mathcal{E} and ρ , and construct the corresponding “basis” set $\mathbb{C} = \{P_i\}$. This step reduces the number of independent parameters. [Sec. IIID]
2. Inspect the entanglement classes of pure states regarding \mathcal{E} and divide the set of pure states into the classes. It improves the convergence of the problem. [Sec. IIIB]
3. Set the number of parallel solvers and define the smoothing function. [Sec. IIIB]
4. Solve the minimax optimization; maximize $G(\Pi)$ using a competitive optimization algorithm. At each Π , the value and the derivative of $G(\Pi)$ are obtained via the parallelization and the smoothing. [Sec. IIIB]

5. Verify the global optimality by evaluating $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$. [Sec. III C] When the solution is far from the global optimum, increase the number of parallel solvers and/or subdivide the entanglement classes. Then retry the optimization.

IV. THREE-QUBIT GHZ ENTANGLEMENT

We apply the minimax optimization to the quantification of the three-qubit GHZ entanglement in example states. The example states are composed of three species: The GHZ state $|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$, the W state $|W\rangle = \frac{1}{\sqrt{3}}(|100\rangle + |010\rangle + |001\rangle)$, and the white noise $I/8$. We reproduce the result of Ref. [16] for the mixtures of two out of the three species, and scrutinize the mixture of the three species. The examples reveal some general properties of the form of the optimal witness operator and of the optimal pure-state decomposition, hence, of the convex structure of general mixed states.

This section consists of the introduction of a measure of three-qubit GHZ mixed-state entanglement in Sec. IV A, how to apply our minimax approach to the measure in Sec. IV B, three examples of quantification of GHZ mixed-state entanglement in Secs. IV C – IV E, and the usefulness of d_{\min} in the numerical quantification in Sec. IV F. It also has Sec. IV G where some general properties of the form of the optimal witness operator are discussed.

A. Extensive three-tangle

The entanglement structure of three qubits is more complicated than that of two qubits, as there are six classes in the former while only two classes (entangled or not) in the latter. Among the six classes, there are two genuine tripartite entanglements, GHZ and W, whose representative states are $|\text{GHZ}\rangle$ and $|W\rangle$ respectively [19]. The GHZ\W class, the set of states with finite GHZ entanglement, comprises the outermost part of the convex structure of three-qubit mixed states [20].

We choose the extensive three-tangle \mathcal{T}_3 [16] as the measure of three-qubit GHZ entanglement. It quantifies the three-qubit GHZ entanglement and generalizes the three-tangle τ_3 [21] to be invariant under stochastic local operations and classical communications (SLOCC) for mixed states [16, 22]. It is defined as $\mathcal{T}_3(|\psi\rangle) = \sqrt{\tau_3(|\psi\rangle)}$ for pure states, and extended for mixed states by the convex roof construction. The name “extensive” is given by the property $\mathcal{T}_3(r\rho) = r\mathcal{T}_3(\rho)$ for positive real r . The optimal witness operator for τ_3 can be also constructed by Eq. (4), but we do not discuss it in the present work.

B. Minimax optimization for \mathcal{T}_3

When one applies the minimax formalism in Eq. (4) to \mathcal{T}_3 , the domain of pure states for the inf problem needs to be carefully considered due to the W class. For a mixture ρ of GHZ- and W-class states, \mathbb{P}_{X_ρ} or \mathbb{R} should include the states in both the classes. However, the set of W-class states has zero volume with respect to pure states, which means that the W-class states are hard to be accessed when a pure state is parametrized to cover the overall pure states. Even more, the directional derivative of $\mathcal{T}_3(|\psi\rangle)$ at the W-class pure states in the direction along the GHZ-class pure states shows divergence, originated from the definition of \mathcal{T}_3 ; $D_{|\psi\rangle}\mathcal{T}_3 = (2\sqrt{\tau_3})^{-1}D_{|\psi\rangle}\tau_3$ diverges when $\mathcal{T}_3 = \tau_3 = 0$ and $D_{|\psi\rangle}\tau_3$ is finite. This divergence obstructs the inf problem to be solved correctly.

To avoid the above problems, we separate the domain of the inf problem into the GHZ\W and W classes, as

$$\begin{aligned}\mathcal{E}(\rho) &= \sup_{\Pi \geq 0} \inf_{\substack{|\psi_G\rangle \in \text{GHZ}\backslash W \\ |\psi_W\rangle \in W}} \text{Tr}((\Pi - \tilde{\mu}I)\rho), \\ \tilde{\mu} &= H(\mu_W, \mu_G; b) \gtrsim \mu = \max\{\mu_G, \mu_W\}, \\ \mu_G &= \text{Tr}(\Pi\pi_{\psi_G}) - \mathcal{E}(|\psi_G\rangle), \\ \mu_W &= \text{Tr}(\Pi\pi_{\psi_W}).\end{aligned}$$

Here we parametrize $|\psi_W\rangle$ and $|\psi_G\rangle$ by using the generalized Schmidt decomposition [23],

$$\begin{aligned}|\psi\rangle &= U_1 \otimes U_2 \otimes U_3 |\Psi\rangle, \\ |\Psi\rangle &= \lambda_0|000\rangle + \lambda_1 e^{i\phi}|100\rangle + \lambda_2|101\rangle + \lambda_3|110\rangle \\ &\quad + \lambda_4|111\rangle,\end{aligned}$$

where U_i 's are local unitary, $\lambda_i \geq 0$, and $\sum_i \lambda_i^2 = 1$. $\mathcal{T}_3(|\psi\rangle) = \mathcal{T}_3(|\Psi\rangle) = 2\lambda_0\lambda_4$ holds for GHZ\W-class states and $\lambda_4 = \phi = 0$ for W-class states.

We mention the optimization settings chosen for the examples below. We parallelize the inf problem with 30 solvers. Every solver in the sup and inf problems runs based on the sequential quadratic programming [24] algorithm, an iterative method for nonlinear optimization.

C. Mixture of $|\text{GHZ}\rangle$ and $I/8$

First our formal approach is applied to the mixed state of $|\text{GHZ}\rangle$ and the white noise $I/8$, $\rho_{\text{GI}}(q) = (1-q)\pi_{\text{GHZ}} + qI/8$ with $q \in (0, 1)$. The entanglement quantification for this state was analytically studied in Ref. [16]. We reproduce the result of Ref. [16] numerically.

The symmetry group for ρ_{GI} consists of (i) the local phase rotation

$$R_L(\alpha, \beta) \equiv (\pi_0 + e^{i\alpha}\pi_1) \otimes (\pi_0 + e^{i\beta}\pi_1) \otimes (\pi_0 + e^{-i(\alpha+\beta)}\pi_1)$$

for the continuous range $\alpha, \beta \in [0, 2\pi)$, (ii) all permutations among three parties, and (iii) 0-1 flip for three parties $\sigma_x^{\otimes 3}$. Here $\pi_0 = |0\rangle\langle 0|$, $\pi_1 = |1\rangle\langle 1|$ and σ_x is a

Pauli matrix. We select the bases of symmetric operator as $P_{i=0,1} = \Sigma_i/\sqrt{2}$ and $P_2 = (I - \Sigma_0)/\sqrt{6}$, where we define $\Sigma_{i=0,1,2,3}$ as

$$\begin{aligned}\Sigma_0 &\equiv \pi_{000} + \pi_{111}, & \Sigma_1 &\equiv |000\rangle\langle 111| + |111\rangle\langle 000|, \\ \Sigma_2 &\equiv -i|000\rangle\langle 111| + i|111\rangle\langle 000|, & \Sigma_3 &\equiv \pi_{000} - \pi_{111},\end{aligned}$$

with $\pi_{000} = |000\rangle\langle 000|$ and $\pi_{111} = |111\rangle\langle 111|$.

We obtain $\mathcal{T}_3(\rho_{\text{GI}}(q \leq q_0)) = 1 - q/q_0$ and $\mathcal{T}_3(\rho_{\text{GI}}(q > q_0)) = 0$ for $q_0 \simeq 0.304$, where the global optimality of $X_{\rho_{\text{GI}}}$ is ensured by $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi)) \sim 10^{-7}$. The optimal witness is also found as

$$X_{\rho_{\text{GI}}} \simeq 4.053P_0 + 1.604P_1 - 3.000I \text{ for } q \leq q_0, \quad (11)$$

and $X_{\rho_{\text{GI}}} = 0$ for $q > 0$. The optimal decomposition for $q < q_0$ is $\rho_{\text{GI}}(q) = \mathcal{T}_3(q)\pi_{\text{GHZ}} + (1 - \mathcal{T}_3(q))\rho_Z$, where ρ_Z is the W-class mixed state

$$\begin{aligned}\rho_Z &= \frac{1}{18} \sum_{\substack{s=\pm \\ m,n=0,1,2}} \pi_{Z_{smn}}, \\ |Z_{smn}\rangle &\equiv \left[\bigotimes_{\nu=m,n,-m-n} \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{2\nu\pi}{3}} \end{pmatrix} \right] |Z_{s00}\rangle, \\ |Z_{+00}\rangle &\equiv \sum_{i,j,k=0,1} a_{i+j+k} |ijk\rangle, \\ |Z_{-00}\rangle &\equiv \sum_{i,j,k=0,1} a_{3-(i+j+k)} |ijk\rangle,\end{aligned}$$

and $(a_0, a_1, a_2, a_3) \simeq (0.7436, -0.1750, -0.2133, 0.4677)$. For the above decomposition, the asymmetric unit of $\mathbb{P}_{X_{\rho_{\text{GI}}}}$ has two elements. We choose them as $|\text{GHZ}\rangle$ and $|Z_{+00}\rangle$, as the other $|Z_{smn}\rangle$'s are obtained from $|Z_{+00}\rangle$ by the elements in the symmetry group.

D. Mixture of $|\text{GHZ}\rangle$ and $|W\rangle$

Next we study $\rho_{\text{GW}}(p) = (1 - p)\pi_{\text{GHZ}} + p\pi_W$ with $p \in (0, 1)$, and numerically confirm the analytic result in Ref. [16]. The form of the optimal witness operator for ρ_{GW} depends on the Hilbert space \mathcal{H} under consideration. It has an exact form in the range of ρ_{GW} , while it has an asymptotic form in the full Hilbert space of three qubits. Both the forms give the same result of $\mathcal{T}_3(\rho_{\text{GW}}(p \leq p_0)) = 1 - p/p_0$, $\mathcal{T}_3(\rho_{\text{GW}}(p > p_0)) = 0$ for $p_0 \simeq 0.3731$, and the optimal pure-state decomposition

$$\rho_{\text{GW}}(p) = \begin{cases} \mathcal{T}_3(p)\pi_{\text{GHZ}} + \frac{1 - \mathcal{T}_3(p)}{3} \sum_{n=0,1,2} \pi_{Z'_n(p_0)}, & p \leq p_0, \\ \frac{1}{3} \sum_{n=0,1,2} \pi_{Z'_n(p)}, & p > p_0, \end{cases}$$

where $|Z'_n(p)\rangle \equiv \sqrt{1-p}|\text{GHZ}\rangle - \sqrt{p}e^{2n\pi i/3}|W\rangle$ are W-class states with $\text{Tr}(X_{\rho_{\text{GW}}}\pi_{Z'_n}) = 0$. Without loss of

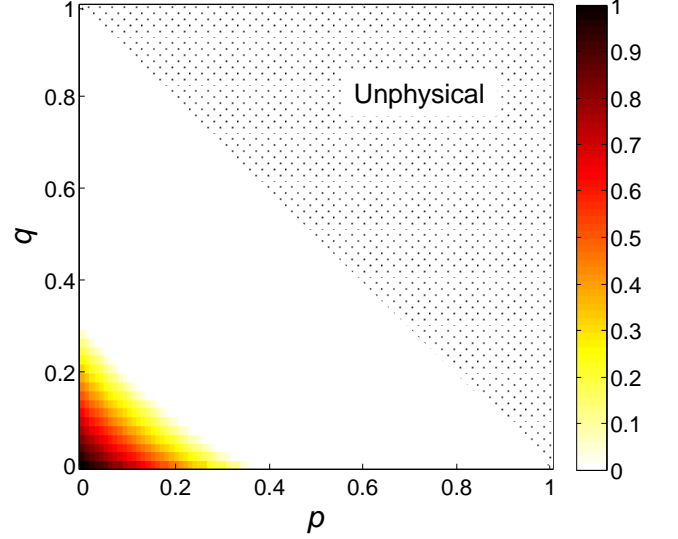


FIG. 1. (Color online) Extensive three-angle \mathcal{T}_3 of the three-qubit full-rank mixed state of $\rho_{\text{GWI}}(p, q) = (1 - p - q)\pi_{\text{GHZ}} + p\pi_W + qI/8$. The physical region of (p, q) is defined by $p, q, (1 - p - q) \in [0, 1]$. The states on the abscissa and on the ordinate are $\rho_{\text{GW}}(p)$ and $\rho_{\text{GI}}(q)$ whose values of \mathcal{T}_3 vanish for $p \geq p_0 \simeq 0.373$ and $q \geq q_0 \simeq 0.304$, respectively.

generality, $[\mathbb{P}_{X_{\rho_{\text{GW}}}}]^A$ consists of $|\text{GHZ}\rangle$ and $|Z'_0(p_0)\rangle$ in the case of $p \leq p_0$, while a single element $|Z'_0(p)\rangle$ for $p > p_0$; the others $|Z'_n(p)\rangle$'s with $n = 1, 2$ are obtained from $|Z'_0(p_0)\rangle$ by the symmetry operator of $R_L(\frac{2n\pi}{3}, \frac{2n\pi}{3})$.

By using this example of ρ_{GW} , we show how the asymptotic form of the optimal witness operator is handled in the minimax formulation. In this case where \mathcal{H} is the full Hilbert space of three qubits, the elements of \mathbb{G} are (i) discrete local phase rotations $R_L(\frac{2n\pi}{3}, \frac{2n\pi}{3})$ with $n = 0, 1, 2$, and (ii) all possible permutations among the three parties. Note that the symmetry group for ρ_{GW} is the subgroup of that for ρ_{GI} . The bases of symmetric operator are

$$\begin{aligned}P_{i=0,1,2,3} &= \Sigma_i/\sqrt{2}, \\ P_4 &= \pi_{W_1}, \quad P_5 = (\pi_{W_2} + \pi_{W_3})/2, \\ P_6 &= \pi_{\bar{W}_1}, \quad P_7 = (\pi_{\bar{W}_2} + \pi_{\bar{W}_3})/2,\end{aligned} \quad (12)$$

where $|W_n\rangle = R_L(0, \frac{2(n-1)\pi}{3})|W\rangle$ and $|\bar{W}_n\rangle = \sigma_x^{\otimes 3}|W_n\rangle$. To tame the divergent parameters of the optimal witness operator of an asymptotic form, we set the bound of the parameters of $\Pi = \sum_i v_i P_i$ as $|v_i| \leq k$ where k is an arbitrarily large positive number. The accuracy of the result increases as k grows; for $k = 10^2$, both of $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ and $\mathcal{T}_3(\rho_{\text{GW}}) - \text{Tr}(X_{\rho_{\text{GW}}}\rho_{\text{GW}})$ are $\sim 10^{-2}$, and for $k = 10^3$ they both decrease to $\sim 10^{-3}$. We obtain, for $p \leq p_0$, that $\text{Tr}(X_{\rho_{\text{GW}}}P_{i=0,1}) \rightarrow 1/\sqrt{2}$, $\text{Tr}(X_{\rho_{\text{GW}}}P_4) \rightarrow 1 - p_0^{-1}$, and the other $\text{Tr}(X_{\rho_{\text{GW}}}P_{i \neq 0,1,4})$'s diverge, as k increases. Regardless of the divergence, $\text{Tr}(\rho_{\text{GW}}X_{\rho_{\text{GW}}})$ successfully converges to the exact value of $\mathcal{E}(\rho_{\text{GW}})$ since $\text{Tr}(\rho_{\text{GW}}P_{i \neq 0,1,4}) = 0$.

E. Mixture of $|\text{GHZ}\rangle$, $|W\rangle$, and $I/8$

We consider a more complicated example of the mixture of $|\text{GHZ}\rangle$, $|W\rangle$ and $I/8$, $\rho_{\text{GWI}}(p, q) = (1 - p - q)\pi_{\text{GHZ}} + p\pi_W + qI/8$ with $p, q, p+q \in (0, 1)$. The symmetry of ρ_{GWI} is the same as that of ρ_{GW} ; see Eq. (12). In Fig. 1, we numerically compute $\mathcal{T}_3(\rho_{\text{GWI}})$ using the minmax optimization. The global optimality of the optimal witness operator is ensured by $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi)) \sim 10^{-4}$ in the computation.

We address the optimal pure-state decomposition of $\rho_{\text{GWI}}(p, q)$ for the case that $\rho_{\text{GWI}}(p, q)$ has finite $\mathcal{T}_3(\rho_{\text{GWI}}(p, q))$. The decomposition form is numerically obtained, after finding the optimal witness operator. It consists of one $\text{GHZ}\backslash W$ -class pure state and multiple W -class pure states as $\rho_{\text{GWI}} = \alpha\pi_r + (1 - \alpha)\rho_Z$, where $\alpha = \alpha(p, q)$, $|r\rangle = \sqrt{1-r}|000\rangle + \sqrt{r}|111\rangle$ is a $\text{GHZ}\backslash W$ -class state, $r = r(p, q) \in (0, 1)$, and ρ_Z is a mixture of W -class states; this decomposition form indicates that \mathcal{T}_3 has the form of $\mathcal{T}_3(\rho_{\text{GWI}}) = \alpha\mathcal{T}_3(|r\rangle) = 2\alpha\sqrt{r(1-r)}$. ρ_Z is decomposed into pure states in the symmetric set whose asymmetric unit has two elements $|x\rangle$ and $|y_+\rangle$,

$$\rho_Z = \beta_1\rho_x + \beta_2\rho_y, \quad \beta_1 + \beta_2 = 1, \quad \beta_1, \beta_2 \geq 0, \quad (13)$$

$$\rho_x = \frac{1}{3} \sum_{n=0}^2 \pi_{x_n}, \quad |x_n\rangle = R_L\left(\frac{2\pi n}{3}, \frac{2\pi n}{3}\right)|x\rangle,$$

$$\rho_y = \frac{1}{6} \sum_{\substack{m=\pm \\ n=0,1,2}} \pi_{y_{mn}}, \quad |y_{\pm n}\rangle = R_L\left(\frac{2\pi n}{3}, \frac{2\pi n}{3}\right)|y_{\pm}\rangle,$$

$$|x\rangle = \sum_{i,j,k=0,1} b_{i+j+k} |ijk\rangle,$$

$$|y_+\rangle = c_0|000\rangle - c_1|W_2\rangle - c_2|\overline{W}_3\rangle + c_3|111\rangle,$$

$$|y_-\rangle = c_0|000\rangle - c_1|W_3\rangle - c_2|\overline{W}_2\rangle + c_3|111\rangle,$$

where b_i 's, c_i 's and β_i 's depend on p and q .

We describe how the optimal decomposition of ρ_{GWI} is related to those of ρ_{GI} and ρ_{GW} . For ρ_{GWI} with finite \mathcal{T}_3 , $[\mathbb{P}_{X_\rho}]^A$ consists of one $\text{GHZ}\backslash W$ -class state $|r\rangle$ and two W -class states $|x\rangle$ and $|y_+\rangle$. On the other hand, $[\mathbb{P}_{X_\rho}]^A$ contains one $\text{GHZ}\backslash W$ -class state $|\text{GHZ}\rangle$ and one W -class state $|Z_{+00}\rangle$ for ρ_{GI} with nonzero \mathcal{T}_3 (i.e. $q \leq q_0$), while one $\text{GHZ}\backslash W$ -class state $|\text{GHZ}\rangle$ and another W -class state $|Z'_0(p)\rangle$ for ρ_{GW} with nonzero \mathcal{T}_3 (i.e. $p \leq p_0$). As $\rho_{\text{GWI}}(p, q)$ continuously approaches to $\rho_{\text{GI}}(q)$ by $p \rightarrow 0$, we see that $|y_+\rangle \rightarrow |Z_{+00}\rangle$, $\beta_1 \rightarrow 0$, and $\beta_2 \rightarrow 1$. And, $|x\rangle \rightarrow |Z'_0\rangle$, $\beta_1 \rightarrow 1$, and $\beta_2 \rightarrow 0$, as $\rho_{\text{GWI}}(p, q \rightarrow 0) \rightarrow \rho_{\text{GW}}(p)$.

This example of $\rho_{\text{GWI}}(p, q)$ is the simplest full-rank mixed-state that is composed of pure states with different genuine multipartite entanglement such as GHZ and W . Entanglement in this type of mixed states has not been quantified before, hence, this example demonstrates the usefulness of our numerical approach. As discussed in the following subsections, this example demonstrates the usefulness of d_{\min} in Sec. IV F and reveal some general properties of the form of the optimal witness operator

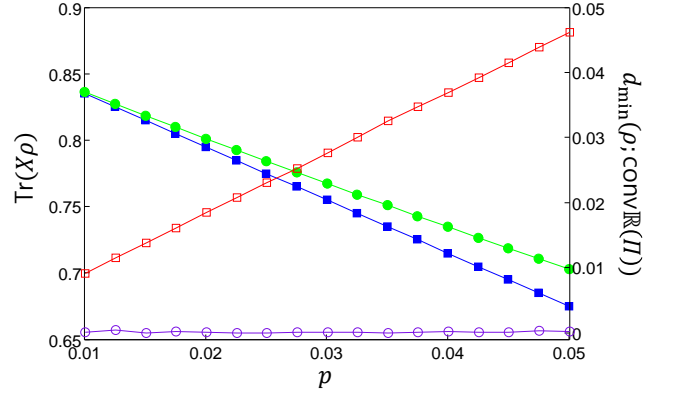


FIG. 2. (Color online) Quantification of entanglement in $\rho(p) = \rho_{\text{GWI}}(p, q = 0.0380)$ by \mathcal{T}_3 . The exact value of $\mathcal{T}_3(\rho)$, $\text{Tr}(X_\rho\rho)$, is numerically obtained by the optimal witness operator X_ρ , and plotted by green filled circles (left ordinate) as a function of $p \in [0.01, 0.05]$. The global optimality of X_ρ is verified by the minimal distance $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi)) \sim 10^{-4}$, which is also plotted by purple open circles (right ordinate). This exact quantification is compared with the underestimation by a misleading less optimal (locally optimal) witness operator X_{les} . The underestimated value of \mathcal{T}_3 by $\text{Tr}(X_{\text{les}}\rho)$ is plotted by blue filled rectangles. We also plot the minimal distance d_{\min} for this underestimation case by red open rectangles, which clearly shows that X_{les} is not globally optimal.

and of the optimal pure-state decomposition, hence, of the convex structure of mixed states in Sec. IV G.

F. Usefulness of d_{\min}

The example in Sec. IV E shows that the minimal distance $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$ is essential for verifying the global optimality. In Fig. 2, the exact computation of \mathcal{T}_3 by the optimal witness operator $X_\rho(p)$ and a slight underestimation by a misleading locally optimal witness operator $X_{\text{les}}(p)$ are compared for $\rho_{\text{GWI}}(p, q = 0.0380)$; the parameters of the two operators are compared in Fig. 3. The global optimality of the two operators are tested by the minimal distances $d_{\min}(\rho; \text{conv } \mathbb{R}(\Pi))$. The values of $\text{Tr}(X_\rho)$ by the two witnesses X_ρ and X_{les} differ by a small amount ($\sim 10^{-4}$) near $p = 0.01$, hence, it is difficult to check the global optimality of a witness operator solely by the expectation value of $\text{Tr}(X_\rho)$. In contrast, d_{\min} is clearly distinguishable by ~ 0.01 between the two witnesses, even in the case that the values of $\text{Tr}(X_\rho)$ are very similar. This demonstrates the usefulness of d_{\min} for distinguishing the globally optimal witness operator from other locally optimal (less optimal) ones, namely for testing the global optimality of a witness operator.

We discuss the appearance of a misleading less optimal witness. When one searches for X_ρ for general ρ and \mathcal{E} , a misleading less optimal witness X_m may frequently appear if the target state ρ neighbors a state ρ_h with high symmetry. In this case, X_m shares the high symmetry

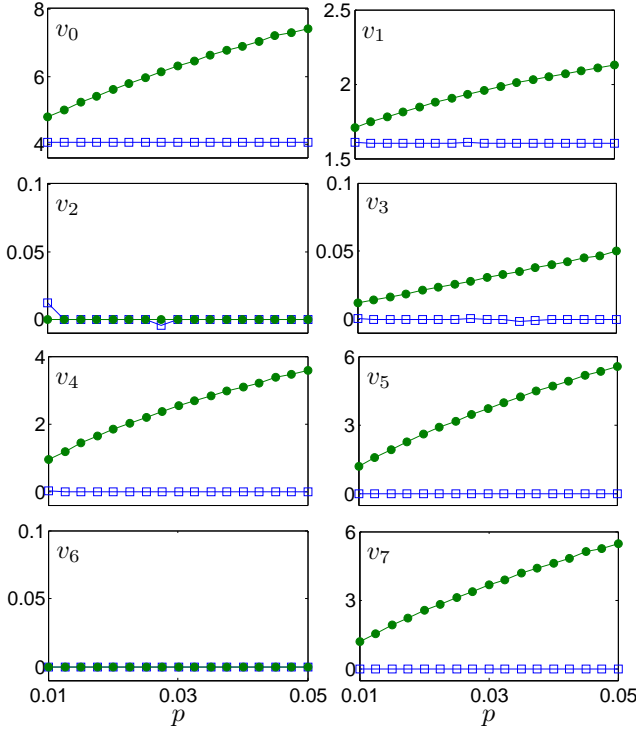


FIG. 3. (Color online) Parameters of $\Pi = \sum_i v_i P_i$ of the two witness operators in Fig. 2, the optimal witness operator X_ρ (green filled circles) and the misleading locally optimal one X_{les} (blue open rectangles), which quantify GHZ entanglement in $\rho_{\text{GWI}}(p, q = 0.0380)$ by \mathcal{T}_3 . We select the bases of symmetric components in Eq. (12), since ρ_{GWI} has the same symmetry as ρ_{GW} .

with ρ_h . Since the high symmetry results in many symmetry operators, \mathbb{P}_{X_m} has lots of elements generated by these operators, and accordingly $\text{conv } \mathbb{P}_{X_m}$ exhibits large dimension and size. This indicates that X_m is a local optimum point to which the optimization easily converges; any small deviation of a witness operator from X_m results in the decrease of $\text{Tr}(X\rho)$ since the deviation degrades the symmetry of witness and reduces $\text{conv } \mathbb{P}_{X_m}$ significantly. In the example of ρ_{GWI} in Sec. IV E, the misleading witness operator X_{les} has the form similar to $X_{\rho_{\text{GI}}}$; one can notice this by comparing the parameters of X_{les} shown in Fig. 3 with Eq. (11). In the parametrization space with respect to Eq. (12), the dimension of $\text{conv } \mathbb{P}_{X_{\rho_{\text{GI}}}}$ is six while $\text{conv } \mathbb{P}_{X_{\rho_{\text{GWI}}}}$ is only three-dimensional. Due to this difference, X_{les} may appear frequently though it is mere local optimum. Such misleading local optimal witnesses can be ruled out by computing d_{\min} .

G. Perturbative construction of optimal witness

When one constructs the optimal witness operator for $\rho' = (1-\delta)\rho + \delta\sigma$ ($\delta \ll 1$), one may guess that the optimal witness operator for ρ is useful, and try to construct the optimal witness operator for ρ' from that for ρ ; we call

this approach the *perturbative construction* of the optimal witness operator. This idea is based upon the continuity of entanglement measure, $|\mathcal{E}(\rho') - \mathcal{E}(\rho)| \rightarrow 0$ as $\delta \rightarrow 0$, and upon the assumption that the form of the optimal witness operator or the form of the optimal pure-state decomposition changes continuously and smoothly from ρ to ρ' . This perturbative approach works well for the states inside each family of ρ_{GI} , ρ_{GW} , and ρ_{GWI} . The success of the perturbative construction saves the number of parallelized solvers for the inf problem in Eq. (4).

However, the perturbative construction is not always successful. For example, one may try to perturbatively construct the optimal witness operator for ρ_{GWI} , starting from that for ρ_{GI} or ρ_{GW} , i.e., from the simpler case for which the optimal witness operator is known. It fails, because of the change of the symmetry (in the case from ρ_{GI}) or the change of the rank (from ρ_{GW}). In the case from ρ_{GI} , the witness for ρ_{GWI} perturbatively constructed from $X_{\rho_{\text{GI}}}$ is X_{les} . As shown in Figs. 2 and 3 (see rectangles), $X_{\text{les}}(p)$ underestimates $\mathcal{T}_3(\rho_{\text{GWI}}(p, q = 0.0380))$, is not globally optimal, and has a form different from $X_\rho(p)$, indicating the failure of the perturbative approach. In the other case from ρ_{GW} , the parameters of the full-rank witness operator for ρ_{GWI} diverge at ρ_{GW} (i.e. at $q = 0$); note that as discussed before, the optimal witness operator for ρ_{GW} has an asymptotic form, when one choose \mathcal{H} as the full Hilbert space of three qubits.

The failure of the perturbative construction is manifested by the non-smooth evolution of the convex structure (with respect to \mathcal{E}) from ρ to ρ' . The non-smooth evolution means that the optimal witness X_ρ or the convex set $\text{conv } \mathbb{P}_{X_\rho}$ evolves non-smoothly between the two cases. We notice this feature, by observing the form of the optimal pure-state decomposition in Secs. IV C – IV E, that \mathbb{P}_{X_ρ} suddenly changes from ρ_{GWI} to ρ_{GI} or ρ_{GW} . In the case of ρ_{GI} , \mathbb{P}_{X_ρ} consists of infinitely many states: One of them is $|\text{GHZ}\rangle$ and the others are the states generated from $|Z_{+00}\rangle$ by the symmetric operators of ρ_{GI} , including the continuous local phase rotation $R_L(\alpha, \beta)$ with $\forall \alpha, \beta \in [0, 2\pi)$. In contrast, there are ten elements $\in \mathbb{P}_{X_\rho}$ in the case of ρ_{GWI} , and only four for ρ_{GW} . Hence, the convex structure of $\text{conv } \mathbb{P}_{X_\rho}$ is non-smoothly connected among ρ_{GWI} , ρ_{GI} , and ρ_{GW} .

This example shows that there is no single universal form of the optimal witness operator in the three-qubit case, in contrast to the two-qubit case [6], and that the convex set $\text{conv } \mathbb{P}_{X_\rho}$ (or its evolution) needs to be studied to construct the optimal witness operator and to understand multipartite or high-dimensional entanglement.

V. CONCLUSION

We develop a numerical optimization approach for finding the optimal witness operator that quantifies mixed-state entanglement. It is applicable to general entanglement measures and states. Importantly, the global optimality of the optimization result is verifiable. Further

development of our approach will provide a powerful tool for studying not only multipartite mixed-state entanglement but also other convexity-related physical properties

such as nonlocality and quantum channels.

ACKNOWLEDGEMENT

We thank NRF for support (Grant No. 2009-0084606).

-
- [1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
 - [2] M. B. Plenio and S. Virmani, *Quant. Inf. Comp.* **7**, 1 (2007); O. Gühne and G. Tóth, *Phys. Rep.* **474**, 1 (2009); F. Mintert, A. R. R. Carvalho, M. Kuś, and A. Buchleitner, *Phys. Rep.* **415**, 207 (2005).
 - [3] M. Horodecki, *Open Syst. Inf. Dyn.* **12**, 231 (2005).
 - [4] W. K. Wootters, *Phys. Rev. Lett.* **80**, 2245 (1998).
 - [5] C. Schmid *et al.*, *Phys. Rev. Lett.* **101**, 260505 (2008); F. Mintert and A. Buchleitner, *ibid.* **98**, 140505 (2007).
 - [6] H. S. Park, S.-S. B. Lee, H. Kim, S.-K. Choi, and H.-S. Sim, *Phys. Rev. Lett.* **105**, 230404 (2010); S.-S. B. Lee and H.-S. Sim, *Phys. Rev. A* **79**, 052336 (2009).
 - [7] B. Röthlisberger, J. Lehmann, D. S. Saraga, P. Traher, and D. Loss, *Phys. Rev. Lett.* **100**, 100502 (2008); B. Röthlisberger, J. Lehmann, and D. Loss, *Phys. Rev. A* **80**, 042301 (2009).
 - [8] A. Streltsov, H. Kampermann, and D. Bruß, *Phys. Rev. A* **84**, 022323 (2011).
 - [9] A. Barvinok, *A Course in Convexity*, American Mathematical Society (2002), p. 10.
 - [10] F. G. S. L. Brandão, *Phys. Rev. A* **72**, 022310 (2005);
 - [11] J. Eisert, F. G. S. L. Brandão, and K. M. R. Audenaert, *New J. Phys.* **9**, 46 (2007).
 - [12] B. M. Terhal, *Theor. Comput. Sci.* **287**, 313-335 (2002)
 - [13] M. Lewenstein, B. Kraus, J.I. Cirac, and P. Horodecki, *Phys. Rev. A* **62**, 052310 (2000)
 - [14] M. Horodecki, P. Horodecki, and R. Horodecki, *Phys. Lett. A* **223**, 1 (1996).
 - [15] O. Gühne, M. Reimpell, and R. F. Werner, *Phys. Rev. Lett.* **98**, 110502 (2007).
 - [16] S.-S. B. Lee and H.-S. Sim, *Phys. Rev. A* **85**, 022325 (2012).
 - [17] V. F. Dem'yanov and V. N. Malozemov, *Introduction to minimax*, Wiley (New York) (1974).
 - [18] B. Rustem, S. Žaković and P. Parpas, *Optim. Method Softw.* **23**, 911 (2008).
 - [19] W. Dür, G. Vidal, and J. I. Cirac, *Phys. Rev. A* **62**, 062314 (2000); C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal, *Phys. Rev. A* **63**, 012307 (2000).
 - [20] A. Acín, D. Bruß, M. Lewenstein, and A. Sanpera, *Phys. Rev. Lett.* **87**, 040401 (2001).
 - [21] V. Coffman, J. Kundu, and W. K. Wootters, *Phys. Rev. A* **61**, 052306 (2000).
 - [22] G. Gour, *Phys. Rev. Lett.* **105**, 190504 (2010).
 - [23] A. Acín *et al.*, *Phys. Rev. Lett.* **85**, 1560 (2000).
 - [24] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd Ed., Springer Series in Operations Research, Springer Verlag (2006).